

Sign-Up or Give-Up: Exploring User Drop-Out in Web Service Registration

Miguel Malheiros
Microsoft Research Cambridge
21 Station Road
Cambridge CB1 2FB, UK
t-mimalh@microsoft.com

Sören Preibusch
Microsoft Research Cambridge
21 Station Road
Cambridge CB1 2FB, UK
spr@microsoft.com

ABSTRACT

Web service providers often require users to create an account to access their services. This registration process usually consists of a web form that users have to fill in with different items of personal data. These data can be leveraged by the provider to create value, but the data requests can make users withdraw from the interaction. Users assess the costs and benefits of the personal data disclosure and if they do not consider the benefits to be larger than the costs they will defect. Past research on user defection on the web has been attitudinal, consisting mostly of questionnaires. We present a large-scale observational study of user drop-out behaviour in web registration forms of three major service providers. We find users are significantly more likely to complete the sign-up process when coming from some pages than others. The implication is that some services – represented by these pages – are considered more valuable than others and worth the effort and privacy cost of registration. We suggest comparing these conversion rates with scores of perceived service value in future work.

Categories and Subject Descriptors

K.4.1 [Computers and Society]: Public Policy Issues—*Privacy*; H.1.2 [Models and Principles]: User/Machine Systems—*Human factors*

General Terms

Human Factors

Keywords

privacy, user behaviour, web forms

1. INTRODUCTION

Most web services require the user to create an account with the service provider. The registration process usually requires the user to fill in a web form with personal data.

This personal data is used by the provider to: (1) create communication channels with the user – through email or phone; (2) create a stable user identity across the service which links all content generated by the user; which, in turn, enables (3) value adding activities such as customer profiling, targeted advertising, or service personalisation.

The registration process creates user friction. The user must spend time and effort filling in these details. Some may be immediate answers the user knows; others will involve a bigger memory effort, collected from documents, or created on the spot [7]. Additionally, the user is asked to pay a privacy cost by providing several items of personal data. Users assess the costs and benefits of complying with data requests [8]. If they consider the value proposition to be positive they will go ahead the disclosure, but if they perceive it to be negative they may withdraw from the service or provide false data. These behaviours result in negative business and data quality impact.

Past research suggests users often decide not to provide personal data to a website or avoid it altogether. In a 2002 survey of German web users by IFAK GmbH & Co. [2] 60% of respondents reported avoiding a website because of privacy concerns. Surveys with American web users report similarly large numbers of personal data disclosure avoidance. The Culnan-Milne survey of 2001 [4] reports 82% have refused to provide personal data to a website and 64% have decided not to use websites because of the data requested. More recently, a Blue Research survey [3] indicates 88% have given incorrect data when registering and 54% may leave if asked to register; and Hodder et al. [6] report 92% have omitted or provides incorrect data to websites.

Personal data disclosure is weighed against the service offered. In a survey of American web users Ackerman et al. [1] found that the attitude towards personal data collection and use depended on the service being offered. For example, 78% of participants would agree to have their behaviour on a web site associated with a persistent identifier to get customised service, but only 44% would agree with the same practice to get cross-site customised advertising. However, these studies are all attitudinal surveys; no observation of actual behaviour was conducted. This is important in privacy research because privacy attitudes have been shown to sometimes differ substantially from privacy behaviour [2].

In this paper we describe an observational study of user drop-out behaviour in web services registration. Using pseudonymised web browsing logs, we investigate the likelihood of a user withdrawing from an interaction with a service provider when asked to register. We compare three main

Copyright is held by the author/owner. Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee.

Symposium on Usable Privacy and Security (SOUPS) 2013, July 24–26, 2013, Newcastle, UK.

providers of online identity and other services: Microsoft, Google, and Yahoo. We focused our analysis on the page a user was visiting before landing in the sign-up page of one of these providers. Our results show that users are more likely to create an account when coming from certain pages than others. This finding contributes to the privacy calculus [5] body of research by empirically testing the assumption that users are more inclined to provide personal data and complete a registration process when they are offered some services than others. This study also constitutes a first attempt to quantitatively describe the phenomenon of user drop-out at the registration hurdle.

In the next section (Section 2) we describe how we analysed user browsing sequences to find instances where the user had landed on a signing-up page. In Section 3 we investigate the likelihood a user will create an account depending on the provider and the page they visited before. We interpret these results in Section 4 and discuss their implications for practice and theory in Section 5.

2. METHOD

2.1 Aims

The goal of this study was to observe and analyse web users browsing behaviour when faced with registration web forms for major web service providers. In particular, we were interested in measuring the likelihood of users not completing a registration process that they had initiated. The motivation for this investigation is that when asked to disclose several items of personal data users may decide that the benefits of registration are not enough to compensate the privacy and effort cost of answering. While the literature suggests this is a common behaviour there is little empirical evidence to support those claims. Additionally, we wanted to determine whether the service a user is trying to reach when presented with the registration page affects her/his behaviour. We use the website a user visited before landing on the registration page as a proxy for intent – e.g. if a user visited mail.yahoo.com before registering we assume that it was that service that incentivised them to register. We expected some pages to lead to higher conversion rates than others implying that the perception a user has of a service affects decision to disclose personal data and complete the registration process. Finally, it was a goal of this research to compare behaviour across different service providers. Our research is exploratory in nature and we discuss some caveats to our approach in Section 5.

2.2 Data Source

The findings in this paper result from the analysis of a sampled subset of more than 2 billion sessions of Internet Explorer users who had given their explicit consent to have their web logs recorded. Data comprised the period of February to April 2013. Sessions begin when the browser is opened and end when it is closed. We focus on sessions as opposed to users since different sessions may correspond to different users sharing the same machine¹. Records for each session contain the URLs visited by the user and timestamps for those visits, providing a chronological sequence of

¹A test analysis where sessions from the same machine were grouped together yielded similar results to the ones presented here.



Figure 1: Registration URL sequence matching. Sequences of websites users visited in a session are compared against the known URL sequence for registering an account with the provider. The matching algorithm determines if: (1) the registration process was initiated; (2) the registration process was initiated and finished.

webpages browsed. Records did not contain personally identifiable data or details in secure connections.

2.3 Analysis

URL sequences corresponding to web browsing sessions were parsed to look for web service registration pages (see Figure 1). We searched for the registration pages of three major web services providers: Microsoft, Google, and Yahoo, which provide online applications for mail, word processing, or storage. If a registration page was present in the session, the session was marked as having started the registration process for that specific provider. The following URLs in the sequence were then compared with the URLs of the pages a user must go through to finish the registration successfully. If the whole sequence had been traversed by the user, the session was marked as having completed the registration process; if the user had abandoned the process, the session was marked as not having completed the registration process. For each session that had initiated a registration process we also recorded the last URL visited before starting the registration process.

3. RESULTS

3.1 Impact of Service Provider on Likelihood to Register

Table 1 shows, for each service provider, the number of sessions where the user started and finished the registration process and the number of sessions where the user started but did not finish the registration process. The total num-

Table 1: Registrations Initiated per Provider

Provider	Sign-Up Completed			Number of Steps
	Yes	No	Odds	
Microsoft	9300	9702	0.96	1
Google	4155	39424	0.11	1 or 3
Yahoo	16776	15635	1.07	3

ber of sessions where registrations were started is 94992. Our first research question was whether the likelihood of a user completing a registration process would be affected by the service provider. The three providers’ registration processes are all based on personal data disclosure using web forms. Microsoft’s requires the user to fill in a one page form; Google’s can be similar, but sometimes requires the user to verify her/his phone number with a one-time token sent to that number; Yahoo’s requires the user to go through three web forms.

For this data, we observed that the percentage of users who, having started the registration process, completed it, did in fact significantly differ by provider: $\chi^2(2, N = 94992) = 19482.70, \phi_c = 0.44, p < 0.001$. There are also significant ($p < 0.001$) different likelihoods of completing the sign-up process when comparing providers one to another. The odds of completing the sign-up process are 9 times higher for Microsoft accounts than Google accounts; 1.1 times higher for Yahoo than for Microsoft; and 10 times higher for Yahoo than Google.

3.2 Impact of Webpage Visited Before on Likelihood to Register

In the second part of our analysis, complete and incomplete registration numbers were grouped by the URL users had visited before initiating the registration process – e.g. registration statistics of all users who had visited mail.yahoo.com before initiating Yahoo registration were pooled together. We also merged these URLs according to their domain to account for national-specific versions of the same site – e.g. registration data of users who visited www.google.com and users who visited www.google.co.uk were merged.

3.2.1 Top Websites

Looking at the 50 websites most commonly visited by users before initiating a sign-up process, there was a significant difference in the likelihood to complete the sign-up depending on where the user had come from. For Microsoft we have $\chi^2(49, N = 13736) = 937.31, \phi_c = 0.26, p < 0.001$. This is also the case for Google account sign-up, $\chi^2(49, N = 37109) = 484.18, \phi_c = 0.11, p < 0.001$. Yahoo sign-up yields similar results: $\chi^2(49, N = 26528) = 2147.47, \phi_c = 0.29, p < 0.001$.²

3.2.2 Specific Services vs. Average

We also wanted to investigate whether some services lead

²Values of ϕ must be considered lower bounds for effect size since the less uniform the marginal distributions in a contingency table, the smaller is ϕ for the same odds ratio; thus, when comparing two different conditions, the odds ratio provides an easier interpretation of the relationship between variables (see Figure 2).

Table 2: Microsoft Registrations by Page (Sample)

Webpage	Sign-Up Completed		
	Yes	No	Odds
officesetup.getmicrosoftkey.com	279	76	3.67
skydrive.live.com	22	9	2.44
xbox.com	65	35	1.86
office.microsoft.com	671	377	1.78
mail.live.com	110	101	1.09
average	9300	9702	0.96
support.microsoft.com	19	24	0.79
windows.microsoft.com	528	737	0.72
privacy.microsoft.com	6	13	0.46
answers.microsoft.com	2	6	0.33
microsoft.com	101	426	0.24

to a higher conversion rate than average for that provider. For example, are users more likely to register for a Microsoft account if they come from the Microsoft Office webpage? The odds of completing the Microsoft sign-up process when coming from office.microsoft.com are 1.88 times higher than the odds of completing the process in the average case, $p < 0.001$ (see Table 2). What if the user already has Microsoft Office and is trying to activate it? In this case the incentive to complete the registration should be higher. The data supports this hypothesis. The odds of completing the Microsoft sign-up process when coming from officesetup.getmicrosoftkey.com are 3.85 times higher than the odds of completing the process in the general case, $p < 0.001$. The odds of completing the Microsoft sign-up process when coming from officesetup.getmicrosoftkey.com are 2 times higher than the odds of completing the process in the office.microsoft.com case, $p < 0.001$. However, for users coming from pages such as windows.microsoft.com or answers.microsoft.com the odds of conversion are lower than on average (respectively 0.75 and 0.25 times lower) suggesting that these pages provide weaker incentives for registration, $p < 0.001$.

Looking at Google accounts (see Table 3), the odds of completing the Google sign-up process when coming from mail.google.com are 1.7 times higher than the odds of completing the process in the average case, $p < 0.001$. On the other hand, Google Plus seems to provide a weaker incentive. The odds of completing the Google sign-up process when coming from plus.google.com are 1.8 times lower than the odds of completing the process in the average case, $p < 0.001$.

For Yahoo, messenger.yahoo.com seems to provide a strong incentive to create a Yahoo account. The odds of completing the Yahoo registration were 1.82 times higher than in the average case, $p < 0.001$. Users coming from mail.yahoo.com were less likely to register: the odds of these users completing the registration were 1.72 times lower than in the average case, $p < 0.001$ (see Table 4).

3.2.3 Comparison of Services Across Providers

Figure 2 displays the ratio between conversion odds when coming from specific pages and the provider’s average conversion odds. Red lines connect services from different providers that can be considered equivalent – e.g.: Microsoft Of-

Table 3: Google Registrations by Page (Sample)

Webpage	Sign-Up Completed		
	Yes	No	Odds
books.google.com	10	37	0.31
drive.google.com	2	8	0.25
docs.google.com	8	35	0.23
mail.google.com	178	981	0.18
support.google.com	147	1015	0.14
average	4155	39424	0.11
news.google.com	1	10	0.10
youtube.com	920	11640	0.08
play.google.com	43	684	0.06
plus.google.com	44	742	0.06
maps.google.com	1	22	0.05

Table 4: Yahoo Registrations by Page (Sample)

Webpage	Sign-Up Completed		
	Yes	No	Odds
messenger.yahoo.com	123	63	1.95
groups.yahoo.com	83	59	1.41
my.yahoo.com	208	163	1.28
fantasysports.yahoo.com	87	71	1.23
help.yahoo.com	172	156	1.10
average	16776	15635	1.07
games.yahoo.com	64	61	1.05
toolbar.yahoo.com	37	36	1.03
answers.yahoo.com	64	73	0.88
mail.yahoo.com	101	426	0.62
maps.yahoo.com	2	9	0.22

Office and Google Docs offer similar types of applications and functionality. Pages with an odds ratio above 1 correspond to services with higher odds of conversion than the providers’ average and pages with an odds ratio below 1 correspond to services with lower odds of conversion than the provider’s average.

The odds ratios of equivalent services do not match perfectly, nor was that expected, but it is interesting to note that the ranks of odds ratios for cloud storage, office suites, mail, and help pages between Microsoft and Google are consistent. Surprisingly, Yahoo’s mail service conversion odds are lower than the provider’s average conversion odds, contrary to what happens with Google and Microsoft.

3.2.4 Search Engine

The likelihood to create a Microsoft account is significantly different depending on which search website users come from: Bing, Google, or Yahoo: $\chi^2(2, N = 3859) = 221.28, \phi_c = 0.24, p < 0.001$. For Google accounts we get $\chi^2(2, N = 15085) = 17.39, \phi_c = 0.03, p < 0.001$. And for Yahoo sign-up $\chi^2(2, N = 3311) = 9.60, \phi_c = 0.05, p < 0.001$. Looking at the odds of conversion by search engine (see Table 5) it is surprising to see that users are more likely to complete the Microsoft registration when coming from Google and the Yahoo registration when coming from Bing. In particular, when comparing Google and Bing for Microsoft reg-

Table 5: Registration Odds after Search

Search Engine	Provider		
	Microsoft	Google	Yahoo
Bing	0.52	0.11	1.19
Google	1.33	0.15	1.09
Yahoo	0.42	0.11	0.88

istrations, we see that the odds of a user completing the sign-up process are 2.5 times higher if they come from Google, $p < 0.001$. Moreover, only for Google are odds of conversion highest when users are coming from its own search engine.

4. DISCUSSION

The likelihood of completing the registration process differs significantly from one service provider to another. While Microsoft and Yahoo conversion odds are similar, Google’s are lower. Part of the difference can be explained by the presence of a phone number verification step in the process which causes a substantial number of users who already committed to creating an account – i.e. they filled in a one page form – to withdraw. It is likely that if the phone call or text message from Google takes too long users will give up registering. It is not clear if this is intentional: it could reduce the number of spam accounts and also keep new genuine accounts to users who are really committed to using them. Another possible reason for the low Google conversion rate evidenced in our data is that part of the users already have accounts and landed on the sign-up page by mistake. By analysing web browsing logs alone it is impossible to determine how often users landed on the registration page unintentionally.

Users are more likely to complete the registration process when coming from some webpages than when coming from others. The page users visit before landing on the sign-up start page may hint at which service users are trying to obtain when registering for an account with one of the providers. When analysing the odds of conversion of users coming from webpages corresponding to known services we see that some have significantly higher or lower odds of conversion than the average for that provider. This suggests that some services are more valuable to users than others. Users are more likely to complete the registration for a Microsoft account after browsing the Microsoft Office website than they are on average. Conversely they are less likely to create an account after visiting the generic Microsoft Windows website. This suggests that these websites and the services they enable or advertise have different value propositions. The more valuable services are more likely to be considered worth the privacy and effort cost of disclosing personal data, while the least valuable will not. The Microsoft Office website allows the user to download and activate this product after creating an account providing an incentive for registering. The Windows website, on the other hand, is a brochure providing information about the features of the operating system and unlikely to incentivise users to register for an account.

Applying the same reasoning, the Google account results indicates that Google Mail has a higher value proposition for users than Google Plus, while the Yahoo account results suggests that users are more willing to register to get Yahoo!

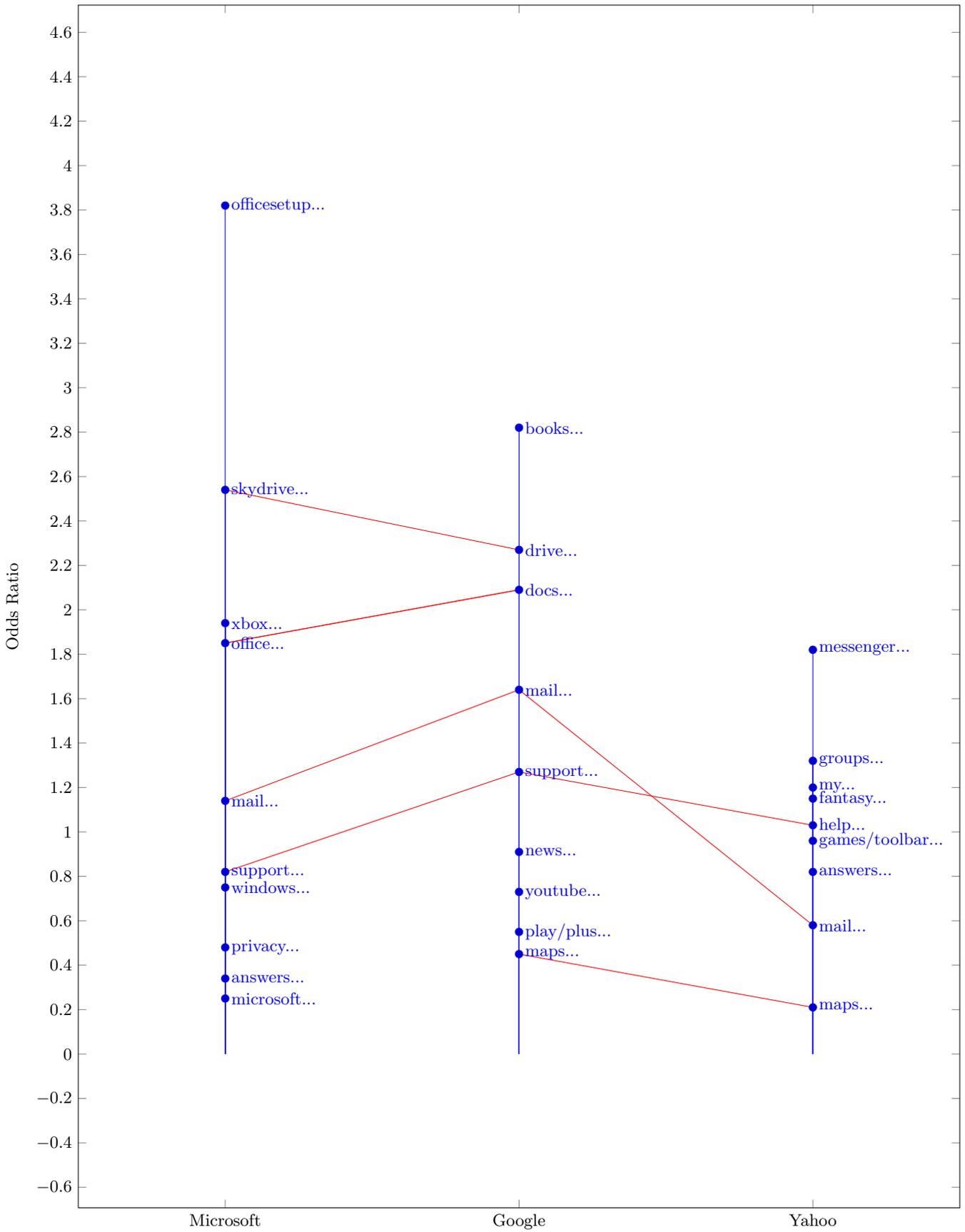


Figure 2: Ratio of conversion odds coming from specific pages and the provider's average conversion odds for a sample of pages.

Messenger than to get Yahoo! Mail. Thus, we propose that the registration conversion rate when coming from different service pages can be used to estimate users' perceptions of service value. However, these estimates have to be validated; one way to accomplish this is to collect ratings of perceived service value from a sample of users and compare them with these conversion odds.

The relative odds ratios of cloud storage, office suite, mail, and support pages were consistent between Microsoft and Google. This result supports the hypothesis that some types of services, once the user is interested in them, are considered more valuable than others, i.e. a user interested in cloud storage is more likely to be willing to pay the effort and privacy cost of registration than one interested in an office suite or mail.

Our results indicate that users are more likely to complete the Microsoft account registration when coming from Google than when coming from Bing. One possibility is that Bing users are more likely than Google users to already have an account with Microsoft and not need to create a new one. Another possible explanation for these results is the presence of a link on Bing's homepage leading to the Microsoft account sign-up page. Users may have clicked the link by mistake or out of curiosity without having the intention to create an account. In this case they were also registered in the logs as coming from the Bing domain. Users coming from the Google domain could only have arrived through a search suggesting intentionality.

5. CONCLUSIONS

We presented a large-scale observational study on user withdrawal behaviour in web service provider registration processes. Analysing user web browsing logs for almost 95 thousand sessions we found that the likelihood of a user registering for an account differs significantly depending on the page visited before initiating the registration. We propose this page gives an indication of the service the user was interested in obtaining and that the conversion rate may be used as a measure of perceived value of the service. We suggest future work can validate this finding by collecting perceptions of service value from a sample of web users and comparing them against these ratings.

This study contributes to the field of privacy calculus by providing empirical evidence that users are more willing to disclose personal data and put in the effort of filling in a form when offered specific services. It also constitutes a first step in quantifying the phenomenon of user drop-out in web registration forms for three major web service providers.

Privacy research is usually focused on users' risk and cost perceptions and less on benefits perception. If a link is shown between certain dimensions of value perception and drop-out behaviour then service providers can work on improving their value proposition when asking for personal data. This would make the interaction fairer from the perspective of the user and would also benefit the provider by increasing conversion rate.

There are some limitations to our approach. We are inferring intention from the webpage the user visited before reaching the registration page. We assume, for example, that if the user visited mail.yahoo.com they were interested in Yahoo's mail service. There are cases, however, where it is very difficult to infer the intention or interest of the user. There are also confounding variables that could realistically

affect the likelihood of a user landing on a registration page. Factors like position of the link to the sign-up page, general layout of the page the user was visiting before, colours used, among others, may have a comparable impact to that of the value of the service the page represents. Users may have landed on the sign-up page by mistake. We assume the likelihood of erroneously clicking on the sign-up link is constant, but this may not be the case. It is also possible that some users may have disclosed false data when registering (falsification of data could be caused by perceptions of unfairness, for example [9]) meaning they are not paying the full privacy cost of signing up and undermining conclusions drawn about the perception of value of the service they were trying to obtain. Finally, some services are available without the need to register. They may be considered very valuable by users and still not lead to a high conversion rate.

The sample has some limitations as well. This study is based on data from Internet Explorer users and, as such, the user sample may be biased towards having preference for Microsoft products or being less technology-literate. Moreover, these users gave their explicit consent for their browsing data to be collected and analysed and as such may have a lower level of privacy concern than the population average.

Future work will focus on gathering perceptions of value for the services present in the logs and comparing them with conversion rates. We also intend to further explore these logs by looking at other types of service providers, such as social networks, e-commerce websites, and news websites.

6. REFERENCES

- [1] M. S. Ackerman, L. F. Cranor, and J. Reagle. Privacy in e-commerce: examining user scenarios and privacy preferences. In *Proceedings of the 1st ACM conference on Electronic commerce, EC '99*, pages 1–8, New York, NY, USA, 1999. ACM.
- [2] B. Berendt, O. Günther, and S. Spiekermann. Privacy in e-commerce: stated preferences vs. actual behavior. *Commun. ACM*, 48(4):101–106, Apr. 2005.
- [3] Blue Research. Consumer perceptions of online registration and social login. Technical report, Janrain, 2011.
- [4] M. J. Culnan. The Culnan-Milne survey on consumers & online privacy notices: Summary of responses. In *Interagency Public Workshop: Get Noticed: Effective Financial Privacy Notices*, pages 47–54, 2001.
- [5] T. Dinev and P. Hart. An extended privacy calculus model for e-commerce transactions. *Information Systems Research*, 17(1):61–80, 2006.
- [6] M. Hodder, E. Churchill, and J. Cobb. Lying and hiding in the name of privacy. Technical report, Customer Commons, 2013.
- [7] C. Jarrett and G. Gaffney. *Forms that Work: Designing Web Forms for Usability*. Interactive Technologies. Elsevier Science, 2009.
- [8] R. S. Laufer and M. Wolfe. Privacy as a concept and a social issue: A multidimensional developmental theory. *Journal of Social Issues*, 33(3):22–42, 1977.
- [9] M. Malheiros, S. Preibusch, and M. A. Sasse. Fairly truthful: The impact of perceived effort, fairness, relevance, and sensitivity on personal data disclosure. In *6th International Conference on Trust & Trustworthy Computing (TRUST 2013)*, London, UK, 2013.